

Rezarta Islamaj Doğan, Ph.D.

Contact Information

8600 Rockville Pike
Building 38A, Room 1003D
Bethesda, MD 20894

Phone: (301) 435-8769
E-mail: rezarta@gmail.com
www.ncbi.nlm.nih.gov/CBBresearch/Fellows/Dogan/

Research Interests

Bioinformatics. Machine learning, data mining and information retrieval. Biomedical literature, clinical data and medical text analysis. Feature construction, entity recognition and relationship extraction. Semantics, inference and knowledge discovery.

Education

UNIVERSITY OF MARYLAND	Ph.D. Computer Science
Ph.D. thesis: "Feature generation and analysis applied to sequence classification for splice-site prediction"	Fall 2007
Advisor: Lise Getoor	
UNIVERSITY OF MARYLAND	M.S. Computer Science
M.S. paper: "Identification of protein-coding regions"	May 2003
Advisor: Lise Getoor	
BOSPHORUS UNIVERSITY, ISTANBUL	B.Sc. Computer Engineering
B.S. thesis: "Three-dimensional representation of amino-acid characteristics"	June 2000
Advisors: Ethem Alpaydın and Uğur Sezerman	
suma cum laude	

Appointments

RESEARCH FELLOW	August, 2008 - present
NIH/NLM/NCBI, Bethesda, MD	
POST-DOCTORAL RESEARCH ASSOCIATE	March, 2008 - July, 2008
Joint appointment at Departments of Cell Biology and Molecular Genetics and Computer Science, University of Maryland at College Park, MD	
PRE-DOCTORAL FELLOW	June, 2002 - December 2007
NCBI Scientific Visitors Program, Bethesda, MD	

Publications

Book Chapters

- [1] "A feature generation algorithm with applications to biological sequence classification" with Lise Getoor and W. John Wilbur, Chapter in *Computational Methods of Feature Selection*, Huan Liu and Hiroshi Motoda editors (2007).

Journal Issues

"Machine Learning for Biomedical Literature Analysis and Text Retrieval" with Lana Yeganova, Special Issue in *BMC Bioinformatics* (May 2011).

Journal Papers

- [1] “A context-blocks model for identifying clinical relationships in patient records” with Aurélie Névéal, and Zhiyong Lu, *BMC Bioinformatics* (May 2011)
- [2] “Topics in machine learning for biomedical literature analysis and text retrieval” with Lana Yeganova, *BMC Bioinformatics* (May 2011)
- [3] “The Protein-Protein Interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text” with Martin Krallinger, Miguel Vazquez, Florian Leitner, David Salgado, Andrew Chatranyamonti, Andrew Winter, Livia Perfetto, Leonardo Briganti, Luana Licata, Marta Iannuccelli, Luisa Castagnoli, Gianni Cesareni, Mike Tyers, Gerold Schneider, Fabio Rinaldi, Robert Leaman, Graciela Gonzalez, Sergio Matos, Sun Kim, W. John Wilbur, Luis Rocha, Ashish V Tendulkar1, Shashank Agarwal, Feifan Liu, Xinglong Wang, Rafal Rak, Keith Noto, Charles Elkan, Zhiyong Lu, Jean-Fred Fontaine, Miguel A. Andrade-Navarro and Alfonso Valencia, *BMC Bioinformatics*(to appear)
- [4] “An evolutionary algorithm approach to feature extraction from sequence data and its application to DNA splice-site prediction” with Uday Kamath, Jack Compton, Kenneth De Jong and Amarda Shehu, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, (May 2011)
- [5] “Semi-automatic semantic annotation of PubMed queries: a study on quality, efficiency, satisfaction” with Aurélie Névéal, and Zhiyong Lu, *Journal of Biomedical Informatics*, (November 2010).
- [6] “Click-words: Learning to predict document keywords from a user perspective” with Zhiyong Lu, *Bioinformatics*, (August 2010).
- [7] “Extracting Rx information from clinical narrative” with James G. Mork, Olivier Bodenreider, Dina Demner-Fushman, François-Michel Lang, Zhiyong Lu, Aurélie Névéal, Lee Peters, Sonya E. Shooshan and Alan R. Aronson, *Journal of the American Medical Informatics Association*, (June 2010).
- [8] “Understanding PubMed® user search behavior through log analysis” with G. Craig Murray, Aurélie Névéal, and Zhiyong Lu, *Database (Oxford)*,.
- [9] “Features generated for computational splice-site prediction correspond to functional elements” with Lise Getoor, W. John Wilbur and Stephen M. Mount, *BMC Bioinformatics*, (October 2007).
- [10] “SplicePort: an interactive splice-site analysis tool” with Lise Getoor, W. John Wilbur and Stephen M. Mount, *Nucleic Acids Research*, (June 2007).
- [11] “Structural footprinting in protein structure comparison: the impact of structural fragments.” with Elena Zotenko, W. John Wilbur, Diane P. O’Leary and Teresa M. Przytycka, *BMC Structural Biology*, (August 2007).

Refereed Conference Proceedings

- [1] “A textual representation scheme for identifying clinical relationships in patient records” with Aurélie Névéal and Zhiyong Lu, *International Conference on Machine Learning and Applications*, (December 2010).
- [2] “Author keywords in Biomedical Journal Articles” with Aurélie Névéal, and Zhiyong Lu, *AMIA Annual Symposium*, (November 2010).
- [3] “Identifying protein-protein interactions in biomedical text articles” with Yi Yang, Aurélie Névéal, Minlie Huang and Zhiyong Lu, *BioCreative III*, (September 2010).
- [4] “Visualizing the weakest links: nucleotides vulnerable to mutations that affect splicing” with Stephen M. Mount and Lise Getoor, *Alternative Splicing-Special Interest Group Meeting at ISMB 2008*, Toronto, Canada (July 2008).
- [5] “Characterizing RNA secondary-structure features and their effects on splice-site prediction” with Lise Getoor and W. John Wilbur, *IEEE ICDM Workshop on Mining and Management of Biological Data*, (October 2007).
- [6] “Feature generation algorithm: with application to splice-site prediction,” with Lise Getoor and W. John Wilbur, *European Conference on Principles and Practice of Knowledge Discovery in Databases*, Berlin, Germany (September 2006).

- [7] “A feature generation algorithm for sequences with application to splice-site prediction” with Lise Getoor and W. John Wilbur, *International Workshop on Feature Selection for Data Mining: Interfacing Machine Learning and Statistics*, Bethesda, Maryland (April 2006).
- [8] “Three dimensional representation of amino acid characteristics” with Uğur Sezerman and Ethem Alpaydın, *23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Istanbul, Turkey (October 2001).

Invited Talks

- [1] “Literature Mining for the biomedical domain” *Theoretical and Applied Bioinformatics lecture*, Foundation for Advanced Education in the Sciences, NIH, MD (May 2011).
- [2] “The ups and downs of building a disease name recognition model” *Second DC-area IR Experts meeting*, Loyola University Maryland, Columbia, MD (May 2011).
- [3] “Biomedical literature mining for knowledge discovery” *Wayne State University*, Detroit, MI (March 2011).
- [4] “Biomedical literature: user search characteristics, article search characteristics and other challenges” (**keynote**) *BIONETICS*, Boston, MA (December 2010).
- [5] “Machine Learning and Bioinformatics: in an essential relationship with mutual benefits” *BioLearn*, Boston, MA (December 2010).
- [6] “Machine learning techniques for biomedical text retrieval in PubMed®” (**tutorial**) with Lana Yeganova, *International Conference on Machine Learning and Applications*, Miami, FL (December 2009).
- [7] “From recognizing biological sequences, to identifying search keywords: A feature generation framework” *GRAND Seminar*, George Mason University, Fairfax, VA (September 2009).
- [8] “Feature analysis for splice-site prediction” *Department of Biology, Massachusetts Institute of Technology*, Boston, MA (May 2008).
- [9] “Feature analysis for splice-site prediction” *Penn Genomics Institute, University of Pennsylvania*, Philadelphia, PA (December 2007).
- [10] “Feature generation analysis and splice-site prediction” *Memorial Sloan-Kettering Cancer Center*, New York, NY (November 2007).
- [11] “Analysis of splicing motifs” *Splicing Regulator Motifs Workshop*, Erlangen, Germany (September 2006).

Presentations

- [1] “Relationship identification model: how do medical concepts relate in patient records?” *NCBI Computational Biology Branch Seminar*, Bethesda, MD (April 2011).
- [2] “Characterizing user search behavior in PubMed®” with G. Craig Murray, Aurelie Neveol, and Zhiyong Lu, *AMIA Annual Symposium*, (November 2010).
- [3] “A hybrid approach for PPI content identification and method detection” *BioCreative III*, Bethesda, MD (September 2010).
- [4] “Towards efficient search tools for biomedical databases: Characterizing user search habits and recognizing their information needs” with G. Craig Murray, Aurelie Neveol, and Zhiyong Lu, *ISMB* (July 2010).
- [5] “Identifying words that users find important for retrieving relevant MEDLINE articles” *NCBI Computational Biology Branch Seminar*, Bethesda, MD (January 2010).
- [6] “Visualizing the weakest links: nucleotides vulnerable to mutations that affect splicing” with Stephen M. Mount and Lise Getoor, *Alternative Splicing-Special Interest Group Meeting at ISMB 2008*, Toronto, Canada (July 2008).
- [7] “The proximity effect on feature construction for splice-site finding” *NIH Graduate Student Research Symposium*, Bethesda, Maryland (May 2006).
- [8] “A machine learning solution for splice-site prediction” *TASSA Annual Conference*, Philadelphia, PA (March 2006).

- [9] “Feature generation for sequences with application to splice-site prediction” *NIH Research Festival*, Bethesda, Maryland (October 2005).
- [10] “Finding acceptor splice sites with AdaBoost” *NIH Second Annual Graduate Student Research Symposium*, Bethesda, Maryland (April 2005).
- [11] “Classification of acceptor splice sites using AdaBoost with decision trees” *16th Annual Genomic Sequencing and Analysis Conference*, Washington DC (September 2004).
- [12] “Identification of internal coding regions in mRNA sequences” *NIH First Annual Graduate Student Research Symposium*, Bethesda, Maryland (April 2004).

Service

PROGRAM COMMITTEE:

BIOSIGNALS International Conference on Bio-inspired Systems and Signal Processing, Portugal, 2012
 ACM SIGHT International Health Informatics Symposium, Miami, Florida, 2012
 ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM BCB) 2011
 International Conference for Machine Learning and Applications, Honolulu, Hawaii, 2011.
 International Conference for Machine Learning and Applications, Washington, DC, 2010.
 BIONETICS, Boston, December 1-3, 2010.
 Workshop on Evolutionary Computation and Machine Learning in Bioinformatics (BioLearn), Boston, December 2, 2010.

PROFESSIONAL REVIEWING:

Database: The Journal of Biological Databases and Curation
 PLoS ONE Journal
 Journal of Bioinformatics and Computational Biology.
 BMC Bioinformatics Journal
 Journal of Biomedical Informatics (JBI).
 Journal of Machine Learning.

ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM BCB) 2011
 AMIA 2011 Annual Symposium, Washington DC USA, 2011.
 International Conference for Machine Learning and Applications, Washington, DC, 2010.
 Bioinformatics track of BIONETICS 2010.
 AMIA 2010 Annual Symposium, Washington DC USA, 2010.
 AMIA 2009 Annual Symposium, San Fransisco, 2009.

ORGANIZER:

Special Session on Machine Learning for Biomedical Literature Analysis and Text Retrieval at the International Conference for Machine Learning and Applications 2011, Hawaii, December 18-21, 2011.
 Special Session on Machine Learning for Biomedical Literature Analysis and Text Retrieval at the International Conference for Machine Learning and Applications 2010, Washington DC, December 12-14, 2010.

Honors and Awards

Conference (ICMLA 2010) Travel Award, IEEE	2010
Conference (ICMLA 2009) Travel Award, IEEE	2009
NIH pre-doctoral Fellow	2002-2007
Faculty Horizons	June 2006
Graduated suma cum laude, Computer Science and Engineering, Bosphorus(Bogazici) University	June 2000
Dean's Honor List	1996-2000
Turkish Prime Ministry Fellowship	1997-2000
Rumeli Foundation Fellowship	1997-2000

Ranked 1st in National Mathematical Olimpiad, Albania
Honorable Mention in 13th Balkan Mathematical Olympiad, Bacau, Romania
37th International Mathematical Olympiad, Mumbai, India

March 1996
April 1996
July 1996